

## СТРУКТУРЫ ДОКУМЕНТАЛЬНОГО ПРОСТРАНСТВА СИСТЕМЫ VEDA

Костенко К.И., Левицкий Б.Е., Назаренко С.В.

Кубанский государственный университет

Краснодар, ул. Ставропольская 149.

[kostenko@kubsu.ru](mailto:kostenko@kubsu.ru)

Работа выполнена при финансовой поддержке РФФИ грант №00-07-96005

Программный комплекс VEDA предназначен для создания полей структурированных и формализованных знаний в форме документных сред, ориентированных на применение технологий обработки знаний, а также автоматического управления жизненными циклами документов [1]. В настоящей работе излагаются элементы подхода по развитию выразительных возможностей документального пространства на основе включения в его структуру компонент, содержащих описания схем классификации и структуризации систем документов и их частей. Рассматриваются задачи, решение которых поддерживается иерархиями информационных структур, создаваемых над системами связанных документов.

Базовая информационная модель документных пространств [2], включает метамодели данных связанных документов и процессов вместе с конкретными объектами и не зависит от специфики предметной области. Система связанных документов и их частей порождает базовую семантическую сеть документального пространства, образующую нижний уровень создаваемой на ее основе иерархии, составленной из семантических метасетей.

Узлы иерархии представляются сетями, задающими структурные и функциональные описания схем факторизации сетей предшествующих уровней. В описаниях формализована многозначная классификация, реализующая различные точки зрения на происхождение, содержание, предназначение и способы применения сведений, содержащихся в документах. Такое многообразие точек зрения естественно возникает тогда, когда система смысловых связей между ними порождает качественно новые возможности при работе с информацией. Описания метасетей позволяют выделять и индексировать информационные ресурсы, соответствующие таким описаниям, организовывать доступ, поиск и навигацию по документальному пространству. Для этого применяется процедура унификации описаний семантических сетей внутренних узлов с фрагментами семантических сетей последующих уровней, которая включает операции сравнения в отношениях “являться” и “быть частью” и оценки степени совпадения.

Выделенный фрагмент базовой семантической сети образует кластер документов, если множество внешних связей с его элементами перестраиваем в систему ссылок, начинающуюся с ссылки на кластер, что позволяет организовывать защиту и управление доступом к информационным ресурсам.

Основными видами классификаций, формируемым на основе метасетей: по вложенности структуры и содержания документов, подобно фрагментам базовой сети регулярным структурам связанных документов и их частей, представляемых параллельно-последовательными сетями, описаниям информационных потребностей пользователей, жизненных циклов и процессов времени существования документов и классов документов в информационной среде.

Структуризация и анализ содержания документов осуществляется с помощью систем правил для этапов начальной структуризации, классификации отдельных частей, установления связей между ними.

Литература

1. Kostenko K.I., Levitskii B.E., Nazarenko S.V. Open information and educational spaces in system VEDA, TELEMATICA'2001, Int. conf. on Telematics and Web-Based Education, St.Petersburg, 2001, p.52-53.

2. Костенко К. И., Назаренко С. В., Левицкий Б.Е. Профессиональные электронные энциклопедии в системе VEDA , Вторая Всероссийская научная конференция “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”, г. Протвино, 2000, с. 199-204.

## STRUCTURES OF DOCUMENTARY SPACE OF THE VEDA SYSTEM

Kostenko K.I., Levitskij B.E., Назаренко S.V.

Kuban the state university

Krasnodar, Stavropolskaya street. 149

[kostenko@kubsu.ru](mailto:kostenko@kubsu.ru)

This work is executed at financial support RFBR the grant № 00-07-96005

Program complex VEDA is intended for creating the fields of formalized knowledge in the form of structured documents storages focused on applications of knowledge processing technology and automatic control of the documents life cycles [1]. This article deals with elements of the approach on expansion of the possibilities of documents spaces for knowledge domains to represent general knowledge properties by means of inclusion in its structures special components that contain description of classification schemes and structures for collections of documents and their parts. Applications are considered for problems which decision is supported by hierarchies of the information structures created above systems of connected documents.

The information model of documentary space [2] includes metamodel for structures of connected documents and processes together with concrete objects and does not depend on a subject domain features. The system of the connected documents and their parts is considered as the base semantic network of documentary space forming foundation of hierarchy of semantic metanets.

Elements of such a hierarchy are represented by networks for specifying structural and functional descriptions of fragmentation and generalization of networks of semantic hierarchy previous levels. These descriptions realize the multiple-valued classification for various points of view on an origin, maintenance and ways of application of the data contained in formalized documents. Such multiplicity of the points of view naturally arises when the system of semantic relations for documents and their parts derivates new quality of opportunities for processing the information.

Metanets descriptions allow to allocate and index the information resources appropriate to such descriptions, to organize access, search and navigation on documentary space. Unification procedure for internal semantic networks descriptions with fragments of semantic networks on the subsequent levels of hierarchy is applied for these purposes. It includes operations of comparison in relations “to be” and “be a part” and estimations of a degree of concurrence.

The allocated fragment of a base semantic network derivates the cluster of documents if the set of external connections with its elements may be replaced by system of connections series beginning with the reference on cluster., that allows organize protection and management for access to information resources. The following types of the classifications are represented by metanets for:

- an enclosure of documents structure and the contents;
- similarity of fragments of a base network to regular structures of the connected documents and their parts represented by parallel-serial networks;
- descriptions of users information needs;
- life time cycles of processes, documents and classes of documents.

Recognition of structure and the analysis of the documents content is performed by rules based knowledge model with stages of initial structure recognition, semantic classifications of separate parts and establishments of communications between them.

## References

1. Kostenko K.I., Levitskii B.E., Nazarenko S.V. Open information and educational spaces in system VEDA, TELEMATICA ' 2001, Int. conf. on Telematics and Web-Based Education, St. Petersburg, 2001, p.52-53.
2. 2. Kostenko K.I., Nazarenko S.V., Levitskii B.E. Professional electronic encyclopedias in system VEDA, the Second All-Russia scientific conference " Electronic libraries: perspective methods and technologies, electronic collections ". Protvino, 2000, p. 199-204.